PAPER

Optimizing Radio Resource Allocation in Multimedia DS-CDMA Systems Based on Utility Functions*

Xiang DUAN[†a)], ***Nonmember***, **Zhisheng NIU**[†b)], ***Member***, ***and*** **Junli ZHENG**[†], ***Nonmember***

SUMMARY This paper addresses the utility-based radio resource allocation problem in DS-CDMA systems carrying multimedia traffic. The proposed scheme, aiming at achieving optimal resource allocation, considers the joint power and data rate allocation. To avoid high computational complexity of nonlinear optimization, we reformulate the radio resource allocation problem as a market model, where resource is regarded as a commodity. Since the market model satisfies the incentive-compatible constraint, the optimal resource allocation can be obtained at the market equilibrium in a distributed manner. According to whether to allocate a minimal transmission data rate to each user, two algorithms, UCA and FCA, are proposed. UCA emphasize on maximizing system overall utilities, while FCA guarantees fairness to users. Simulation results show that the proposed radio resource allocation scheme and algorithms are flexible and efficient for multimedia DS-CDMA systems.

key words: *radio resource allocation, DS-CDMA, utility, optimization, market, equilibrium*

1. Introduction

In existing and emerging communication networks, it is of increasing demands to carry a broad spectrum of multimedia traffic with diverse quality-of-service (QoS) requirements. In bandwidth limited wireless systems where channel conditions are time-varying and location-dependent, radio resource allocation is one of the essential technologies for providing QoS to users [1]. The task of radio resource allocation is to control the behavior of user terminal via allocating proper quota of radio resources to users, so that the desired results, such as QoS provisioning, fairness, or efficient resource utilization, can be obtained.

Although most previous studies have interpreted user's QoS requirements into specific technical measurements such as bandwidth, delay and loss probability [2], [3], QoS in fact is the perception of users or upper layer applications [4]. Based on this consideration, some researchers have proposed *utility*-based QoS frameworks in communication networks [5]–[12]. Utility, which in economics describes the welfare a consumer benefits when he or she buy a commodity or service [13], represents the level of satisfaction that a user or an upper layer application derives when using networks services. By mapping the QoS provided by networks as utility, which is interpreted as a "soft" object, the network service is more flexible and therefore tunable according to system load and channel conditions. However, since most utility functions are nonlinear, the utility-based radio resource allocation problems are often non-trivial.

In direct-sequence code-division multiple access (DS-CDMA) systems, power control is the key component of radio resource allocation. In [5] and [6], authors proposed distributed utility-based power control schemes for uplink and downlink in CDMA systems, respectively. Reference [7] proposed a distributed power control scheme based on a N-person non-cooperative game model, and [8] improved the scheme by introducing cost to utility functions. Authors in [9] and [10] considered different utility-based game-theoretical uplink power control models for CDMA systems, and present different approaches to evaluate performance.

In these references, however, only transmitted power is regarded as the controllable resource, and QoS provided is evaluated by the received signal-to-interference ratio (SIR). As we know, in DS-CDMA systems, transmission data rates are adjustable by regulating the variable spreading gains (VSG) of users [14]. Therefore, data rate allocation should also be regarded as a part of radio resource allocation mechanism, which is not considered in [5]–[10]. In [11], the author presented a utility-based power and rate joint control control scheme for CDMA networks which tries to enhance the resource utilization using congestion price. However, the scheme was based on incentive of individual users, which was not able to get the social optimum. A revenue-suboptimal resource allocation scheme for HDR downlink is presented in [12], where resource is allocated based on each user's bidding price. Unfortunately, the authors did not consider the power control issue. Moreover, there are some stringent constraints on users utility functions, which may not be feasible for multimedia traffic with diverse QoS requirements.

As we know, future wireless networks will carry traffic that are highly asymmetric, requiring much more bandwidth in the downlink than in the uplink. This leads to the growing importance of radio resource allocation in downlink. Therefore, in this paper we investigate the utility-based radio resource allocation problem for multimedia DS-CDMA downlink, considering both transmitted power and data rate allocation. The radio resource allocation problem is formulated as an optimization problem, which maximizing the system overall utilities (defined as the sum of utilities

Manuscript received December 9, 2003.
Manuscript revised March 17, 2004.

[†]The authors are with the State Key Lab on Microwave and Digital Communications, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.
a) E-mail: xiang99@mails.tsinghua.edu.cn
b) E-mail: niuzhs@tsinghua.edu.cn

*This work has been supported by the National Science Foundation of China (60272021).

of all users in the system). Then the nonlinear optimization problem is reformulated as market model for reducing computational complexity. Through the adjustment of resource price, the market converges to the its equilibrium, which is the optimal resource allocation. According to whether to allocate each user with a minimal transmission data rate, two different algorithms—utility-centric allocation (UCA) and fairness-centric allocation (FCA)—are proposed.

The rest of the paper is organized as follows. In Sect. 2, we present the system model of the utility-based resource allocation problem. Then the market model is described in Sect. 3, where the problem is decomposed into two subproblems and the optimal resource allocation is achieved in a distributed manner. In Sect. 4, a detailed description of the proposed resource allocation algorithms are given. Also the implementation structure of the utility-based radio resource allocation framework is discussed. Section 5 presents the numerical results. Finally we conclude the paper in Sect. 6.

2. System Model of Utility-Based Radio Resource Allocation

We consider a typical cell in a multimedia DS-CDMA system with N active users. Denoted by R_i the transmission rate, by P_i the transmitted power, and by h_i the channel gain of base station to user i. For simplicity, in theoretical analysis we assume that the channel prediction or estimation is perfect. W represents the system bandwidth, and I_i, the background interference at the location of user i. Then, the received signal-to-interference ratio (SIR) of user i is given by

$$\gamma_i = \frac{W}{R_i}\frac{h_i P_i}{\theta_d h_i(\sum_{j\neq i} P_j) + I_i} \tag{1}$$

where θ_d is the downlink orthogonality factor with the typical values falling in the range of $(0.1, 0.6)$ [15].

We assume that a user evaluates the received QoS by its throughput [4]. In wireless environment, the net throughput of user i is given by its achieved *effective bandwidth*, which is the product of transmission rate R_i and the transmission efficiency, described as

$$s_i = R_i E_s(\gamma_i) \tag{2}$$

where $E_s(\cdot)$ is the transmission efficiency function of user i. Transmission efficiency is defined as the percentage of successfully transmitted information bits to overall bits sent [16]. It is a function of received SIR, γ_i. Readers are referred to [16] for detailed information about transmission efficiency.

The utility function of user i is given by $U_i(s_i) = U_i(R_i E_s(\gamma_i))$. The form of $U_i(\cdot)$ should be properly selected in order to reflect the nature of satisfaction level. References [5], [6] has pointed out that utility should be a nondecreasing function satisfying $U_i(0) \approx 0$ and bounded above by $U_{i\max}$, which is the maximal achievable utility of user i. Generally, the form of utility functions of voice, data and

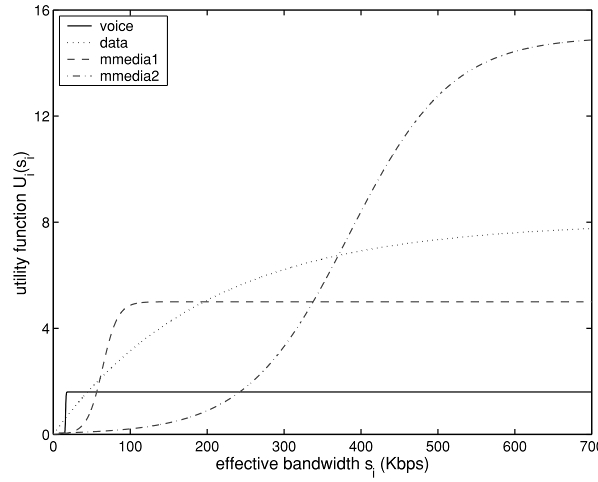

Fig. 1 Utility functions of typical voice, data and multimedia traffic.

multimedia applications are different, as shown in Fig. 1. For voice traffic, a step function can describe its hard QoS requirements on throughput. Since data traffic has no stringent demand on rate, and more throughput leads to higher satisfaction level, its utility is represented by an increasing concave function. The QoS requirements of multimedia traffic are between voice and data, therefore an S-shaped utility function (such as Sigmoid function [5]) is used. Note that for convenience of mathematical treatments, a steep S-shaped function instead of step function are used for voice traffic. However, it should be clarified that our paper is suitable for a wide range of utility functions, not only limited to the examples given in the figure.

Assume that $P_{\max}$ is the transmitted power constraint of base station, and let $\mathbf{R} = \{R_1, R_2, \cdots, R_N\}$ and $\mathbf{P} = \{P_1, P_2, \cdots, P_N\}$. Mathematically, utility-based resource allocation problem can be formulated as an optimization model for maximizing the system overall utilities:

$$\max_{\mathbf{P},\mathbf{R}} \sum_{i=1}^{N} U_i(R_i E_s(\gamma_i))$$
$$\textbf{s.t.} \quad \sum_{i=1}^{N} P_i \le P_{\max} \tag{3}$$

The optimum solution $(\mathbf{P}^*, \mathbf{R}^*)$ to Eq. (3) is called *social optimal resource allocation* for the utility-based radio resource allocation problem.

Equation (3) is a nonlinear optimization problem with $2N$ decision variables. We have the following two theorems for simplifying the system model.

Theorem 1: Assume that $(\mathbf{P}^*, \mathbf{R}^*)$ is the optimum solution to problem (3), then $(\mathbf{P}^*, \mathbf{R}^*)$ satisfies

$$\gamma_i^* = \gamma^*, \quad (i = 1, 2, \cdots, N) \tag{4}$$

where γ^* is the optimal received SIR which is given by

$$\gamma^* = \arg\max_{\gamma} \frac{E_s(\gamma)}{\gamma} \tag{5}$$

proof: We prove this theorem by contradiction. Suppose the assumption in the theorem is not true, i.e., there exists $(\mathbf{P}^*, \mathbf{R}^*)$ that are the optimum solution to Eq. (3), and for some i,

$$\gamma_i^* = \frac{W}{R_i^*}\frac{h_i P_i^*}{I_{-i}^*} \neq \gamma^* \tag{6}$$

where $I_{-i}^* = \theta_d h_i(\sum_{j \neq i} P_j^*) + I_i^*$ is the interference received by user i. It is clear that I_{-i}^* keeps unchanged when data rate changes. Therefore, we let $\mathbf{P}' = \mathbf{P}^*$, $\mathbf{R}' = \{R_1^*, R_2^*, \cdots, R_i', \cdots, R_N^*\}$, where

$$R_i' = \frac{W}{\gamma^*}\frac{h_i P_i^*}{I_{-i}^*} \tag{7}$$

Therefore, $\gamma_i' = \gamma^*$. Since γ^* maximizes $E_s(\gamma)/\gamma$, we get

$$\begin{aligned} R_i' E_s(\gamma_i') &= \frac{W h_i P_i^*}{I_{-i}^*}\frac{E_s(\gamma^*)}{\gamma^*} \\ &> \frac{W h_i P_i^*}{I_{-i}^*}\frac{E_s(\gamma_i^*)}{\gamma_i^*} = R_i^* E_s(\gamma_i^*) \end{aligned} \tag{8}$$

while $R_j^* E_s(\gamma_j)^*$ keep unchanged for $j \neq i$. As a result, $\sum_{i=1}^N U_i(R_i' E_s(\gamma_i')) > \sum_{i=1}^N U_i(R_i^* E_s(\gamma_i^*))$. This contradicts with the fact that $(\mathbf{P}^*, \mathbf{R}^*)$ are the optimum solution to system model (3). So the assumption in the above theorem must be true. □

Theorem 2: Provided that $\gamma_i = \gamma^*$ for all $1 \leq i \leq N$, the constraint conditions

$$\sum_{i=1}^N P_i \leq P_{\max} \tag{9}$$

are equivalent to

$$\sum_{i=1}^N \delta_i g(R_i) \leq 1 \tag{10}$$

where

$$\delta_i = 1 + \frac{I_i^o}{P_{\max}}, \quad I_i^o = \frac{I_i}{\theta_d h_i} \tag{11}$$

and

$$g(R_i) = \frac{1}{\frac{W}{\theta_d R_i \gamma^*} + 1} \tag{12}$$

In this theorem, δ_i is the channel condition factor, and I_i^o is the *equivalent background interference* [17]. We have $\delta_i > 1$, and larger δ_i means worth channel conditions. $g(R_i)$ is called the *power index* [3], which is the measure of resources obtained by a user. $\delta_i g(R_i)$ is the system resources consumption if user i is allocated with data rate R_i [17].

proof: We define

$$\begin{aligned} Z_0 &= \left\{\mathbf{P}, \mathbf{R} | \gamma_i = \gamma^*, \sum_{i=1}^N P_i \leq P_{\max}\right\} \\ Z_c &= \left\{\mathbf{P}, \mathbf{R} | \sum_{i=1}^N \delta_i g(R_i) \leq 1\right\} \end{aligned} \tag{13}$$

Now the problem is converted to the proof of $Z_0 = Z_c$. Assuming that $(\mathbf{P}, \mathbf{R}) \in Z_0$, by mathematical manipulations from Eq. (1), we get

$$P_i = g(R_i)\left(\sum_{j=1}^N P_j + I_i^o\right) \tag{14}$$

or

$$\sum_{i=1}^N P_i = \frac{\sum_{i=1}^N g(R_i) I_i^o}{1 - \sum_{i=1}^N g(R_i)} \tag{15}$$

Since $\sum_{i=1}^N P_i \leq P_{\max}$, we have

$$\sum_{i=1}^N g(R_i) \leq 1 - \sum_{i=1}^N \frac{g(R_i) I_i^o}{P_{\max}} \tag{16}$$

i.e., $\sum_{i=1}^N \delta_i g(R_i) \leq 1$. Therefore, for any $(\mathbf{P}, \mathbf{R}) \in Z_0$, $(\mathbf{P}, \mathbf{R}) \in Z_c$.

On the other hand, if $(\mathbf{P}, \mathbf{R}) \in Z_c$, i.e., they satisfy $\sum_{i=1}^N \delta_i g(R_i) \leq 1$. By manipulation we get

$$\frac{\sum_{i=1}^N g(R_i) I_i^o}{1 - \sum_{i=1}^N g(R_i)} \leq P_{max} \tag{17}$$

Considering Eq. (15), we get $\sum_{i=1}^N \leq P_{\max}$. Thus, $(\mathbf{P}, \mathbf{R}) \in Z_0$. Therefore, $Z_0 = Z_c$. □

Theorem 1 illustrates that the optimal solution $(\mathbf{P}^*, \mathbf{R}^*)$ of problem (3) are not independent; they are coherent by γ^*. Note that the value of γ^*, only determined by the form of transmission efficiency function $E_s(\gamma_i)$, can be obtained by solving Eq. (5) with any one-dimensional optimum search algorithm [18], simulation, or measurements in real systems.

With further consideration of Theorem 2, we get that problem (3) is equivalent to

$$\begin{aligned} &\max_{\mathbf{R}} \sum_{i=1}^N U_i(R_i E_s^*) \\ &\textbf{s.t.} \quad \sum_{i=1}^N \delta_i g(R_i) \leq 1 \end{aligned} \tag{18}$$

where $E_s^* = E_s(\gamma^*)$ is the corresponding transmission efficiency of γ^*.

The equivalent system model (18) has only N independent decision variables, $\mathbf{R}$, which is less than $2N$ in the original system model (3). However, generally finding optimum solution by normal algorithms, such as steepest decent method or gradient projection method [18], is still non-trivial and may be impractical for real systems. In the next section we solve this problem by formulating it into a market model, where the optimal resource allocation can be achieved with lower computational complexity.

3. Market Model: Achieving the Social Optimal Allocation in a Distributed Manner

3.1 Description of Market Model

To solve the problem (18), we first introduce the following lemma given in [18].

Lemma 1: Let $f : \mathbb{R}^N \mapsto \mathbb{R}$ and $h : \mathbb{R}^N \mapsto \mathbb{R}$ be arbitrary functions. Let

$$\mathcal{L}(\mathbf{X}, \lambda) = f(\mathbf{X}) - \lambda h(\mathbf{X}) \tag{19}$$

be the Lagrangian function and

$$\hat{\mathbf{X}}(\lambda) = \arg\max_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \lambda) \tag{20}$$

where $\mathbf{X} = \{x_1, x_2, \cdots, x_N\}$. Then $\hat{\mathbf{X}}(\lambda)$ is a global optimum solution of the following optimization problem:

$$\begin{aligned} &\max_{\mathbf{X}} f(\mathbf{X}) \\ &\mathbf{s.t.}\ \ h(\mathbf{X}) \le h(\hat{\mathbf{X}}(\lambda)) \end{aligned} \tag{21}$$

To make use of this lemma, we define the Lagrangian function of (18) as

$$\begin{aligned} \mathcal{L}(\mathbf{R}, \lambda) &= \sum_{i=1}^{N} U_i(R_i E_s^*) - \lambda\left(\sum_{i=1}^{N} \delta_i g(R_i) - 1\right) \\ &= \sum_{i=1}^{N} \{U_i(R_i E_s^*) - \lambda \delta_i g(R_i)\} + \lambda \end{aligned} \tag{22}$$

and let

$$\hat{\mathbf{R}}(\lambda) = \arg\max_{\mathbf{R}} \mathcal{L}(\mathbf{R}, \lambda) \tag{23}$$

where $\hat{\mathbf{R}}(\lambda) = [\hat{R_1}(\lambda), \hat{R_2}(\lambda), \cdots, \hat{R_N}(\lambda)]$. Thus, according Lemma 1, $\hat{\mathbf{R}}(\lambda)$ is the global optimum solution to the following problem.

$$\begin{aligned} &\max_{\mathbf{R}} \sum_{i=1}^{N} U_i(R_i E_s^*) \\ &\mathbf{s.t.}\ \ \sum_{i=1}^{N} \delta_i g(R_i) \le \sum_{i=1}^{N} \delta_i \hat{g}_i(\lambda) \end{aligned} \tag{24}$$

where $\hat{g}_i(\lambda) = g(\hat{R}_i(\lambda))$. Compare (24) with Eq. (18), we find that if there is a λ^* such that $\sum_{i=1}^{N} \delta_i \hat{g}_i(\lambda^*) = 1$, the global optimum solution can be obtained by $\mathbf{R}^* = \hat{\mathbf{R}}(\lambda^*)$. If $\sum_{i=1}^{N} \delta_i \hat{g}_i(\lambda^*)$ is close to 1 enough, we also find a good approximation to the optimum solution.

Furthermore, since $\mathcal{L}(\mathbf{R}, \lambda)$ is decomposable in $\mathbf{R}$ (Eq. (22)), by defining the revenue function as

$$J_i(R_i, \lambda) = U_i(R_i E_s^*) - \lambda \delta_i g(R_i) \tag{25}$$

we get that $\hat{\mathbf{R}}(\lambda)$ is the solution to (23) if and only if it solves

$$\hat{R}_i(\lambda) = \arg\max_{R_i} J_i(R_i, \lambda),\ (i = 1, 2, \cdots, N) \tag{26}$$

which means the Lagrange maximization problem (23) has been decomposed into N subproblems. Therefore it is possible for solving the problem in a distributed manner.

According to discussions above, in order to achieve the optimum solution to (18), we need to solve its dual problem as

$$\begin{aligned} &\min_{\lambda} \left| 1 - \sum_{i=1}^{N} \delta_i \hat{g}_i(\lambda) \right| \\ &\mathbf{s.t.}\ \ \hat{R}_i(\lambda) = \arg\max_{R_i} J_i(R_i, \lambda), \\ &\qquad\qquad (i = 1, 2, \cdots, N) \end{aligned} \tag{27}$$

In fact, from (3) to (27), we have reformulated the utility-based radio resource allocation problem as a market model. Here radio resource is regarded as a commodity to be exchanged in the market, and λ is the price of resource. Firstly take a further look at the Lagrangian maximization problem, which is decomposed into N independent problems (Eq. (26)). Since λ is the resource price, and $\delta_i g(R_i)$ is the amount of resources consumed by user i, $\lambda \delta_i g(R_i)$ is the expense of user i on using radio resources. Note that $U_i(R_i E_s^*)$ is the achieved utility, therefore $J_i(R_i, \lambda) = U_i(R_i E_s^*) - \lambda \delta_i g(R_i)$ actually represents the revenue (utility minus expense) of user i if allocated with R_i data rate and the current resource price is λ. Hence, the constraint conditions in (27) is interpreted as the revenue maximization problem for each user.

Furthermore, we observe the optimization objective of (27). In fact, in the formula $|1 - \sum_{i=1}^{N} \delta_i \hat{g}_i(\lambda)|$, "1" is the overall supply of radio resource provided by supplier—the network, and "$\sum_{i=1}^{N} \delta_i \hat{g}_i(\lambda)$" is the demand of resources, consumed by users. According to [13], if the demand meet supply in a market with a price, it is called the *equilibrium*, and the price is the *equilibrium price*. Therefore, λ^* is the equilibrium price of the market model. Since that the constraint conditions are revenue maximization problems, the market model denoted by (27) complies with the *incentive-compatibility constraint* [13], where the revenue maximization of individual user coincides with market equilibrium. According to economics theory, with the incentive-compatibility constraint satisfied, the social optimal resource allocation can be achieved at the market equilibrium.

3.2 Finding Social Optimum Allocation: Two Subproblems

As discussed above, the utility-based radio resource allocation problem is reformulated as a market model, which is decomposed into two subproblems:

1. *Revenue maximization*: Each individual user maximizes its own revenue (utility minus expense). This subproblem is to find $\hat{R}_i(\lambda)$ that solve Eq. (26);

2. *Market equilibrium*: The market converges to the equilibrium where supplies equal demands. This subproblem is to look for such resource price λ^* that $\sum_{i=1}^{N} \delta_i \hat{g}_i(\lambda^*) = 1$.

First we consider the revenue maximization subproblem. According to [18], if $J_i(R_i, \lambda)$ is a concave function of R_i, the unique solution to $\max_{R_i} J_i(R_i, \lambda)$ satisfies the first order condition, i.e.,

$$\frac{\partial J(R_i, \lambda)}{\partial R_i} = 0 \tag{28}$$

or

$$E_s^* \frac{\mathrm{d}U_i(R_i E_s^*)}{\mathrm{d}R_i} = \lambda \delta_i \frac{\mathrm{d}g(R_i)}{\mathrm{d}R_i} \tag{29}$$

With the definition of *characteristic function* as

$$f_{\lambda i}(R_i) = E_s^* \frac{\mathrm{d}U_i(E_s^* R_i)/\mathrm{d}R_i}{\mathrm{d}g(R_i)/\mathrm{d}R_i} \tag{30}$$

the first order condition (29) can be rewritten as

$$\delta_i \lambda = f_{\lambda i}(R_i) \tag{31}$$

or equivalently

$$R_i = f_{\lambda i}^{-1}(\delta_i \lambda) \tag{32}$$

Figure 1 shows the characteristic functions of traffic given in Fig. 2. We see that the $f_{\lambda i}(R_i)$ is of a single-peak shape[†]. Let

$$\begin{aligned} \lambda_{i\max} &= \max_{R_i} f_{\lambda i}(R_i) \\ r_{i\min} &= \arg\max_{R_i} f_{\lambda i}(R_i) \end{aligned} \tag{33}$$

It can be proven that $J_i(R_i, \lambda)$ is convex in $[0, r_{i\min}]$ and concave in $[r_{i\min}, +\infty)$. Since the first order condition is applicable where $J_i(R_i, \lambda)$ is concave, Eq. (32) solves problem (26) only when $\delta_i \lambda \leq \lambda_{i\max}$. Therefore, the solution to (26) is

$$R_i = \begin{cases} f_{\lambda i}^{-1}(\delta_i \lambda), & \delta_i \lambda \leq \lambda_{i\max} \\ 0, & \delta_i \lambda > \lambda_{i\max} \end{cases} \tag{34}$$

We call $\lambda_{i\max}$ the *reservation price* of user i, which means the highest tolerable resource price that user i can afford to consume the commodity. $r_{i\min}$ and $\lambda_{i\max}$ of the corresponding types of traffic are also given in Fig. 2.

In Eq. (34), we observe that, if the channel condition of a user is not good (with large δ_i), or the resource price λ rises due to increase of system load, some users cannot afford to consume radio resources. As a result, the portion of users will be allocated with zero transmission data rate, as shown in Eq. (34). Obviously this approach is somewhat unfair for users far away from base station or in a very deep shadowing environment.

In order to avoid this case, we can use another approach to solve Eq. (26) by

$$R_i = \begin{cases} f_{\lambda i}^{-1}(\delta_i \lambda), & \delta_i \lambda \leq \lambda_{i\max} \\ r_{i\min}, & \delta_i \lambda > \lambda_{i\max} \end{cases} \tag{35}$$

Consequently, we propose two resource allocation algorithms according to the two approaches discussed above, which have different targets. One proposal is called *Utility-Centric Allocation* (UCA) algorithm, solving Eq. (26) by Eq. (34), which emphasizes on maximizing overall utilities by each user maximizing its revenue and system balancing the supplies and demands of resource. The other proposal is *Fairness-Centric Allocation* (FCA) algorithm, as Eq. (35) illustrated. The major difference between FCA and UCA is, for any user i in UCA, when current weighted price $\delta_i \lambda$ exceeds its reservation price $\lambda_{i\max}$, the user is simply cut off ($R_i = 0$); while in FCA, the user always obtains a minimal portion of resource allocation by guaranteeing $r_{i\min}$. Since FCA concerns the fairness issue of users while pursuing the system-wide optimum, it only achieves the social suboptimal resource allocation.

Note that Eq. (35) indicates that there is always a minimal data rate $r_{i\min}$ allocated to each user. Therefore, it is possible that $\sum_i \delta_i g(r_{i\min}) > 1$, which means that the system resource cannot meet the total minimum data rate requirements of all users. This case is called *system outage*, which should be avoided or restrained by the call admission control (CAC). This is out of range of this paper and will not be discussed here further.

Then we consider the market equilibrium subproblem. The objective of this subproblem is to adjust the resource price λ to λ^* so that $\sum_{i=1}^{N} \delta_i \hat{g}_i(\lambda^*) \approx 1$. Mathematically, it can be solved asymptotically by the golden section search [18] (or called 0.618-method).

Since the revenue maximization subproblem is individual user based, the social optimal allocation can be achieved

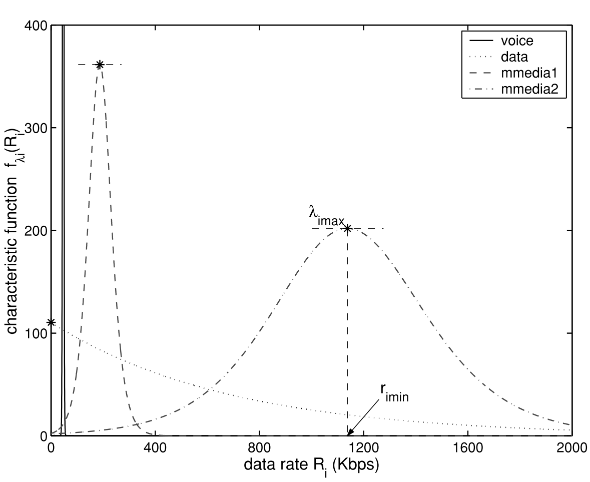

Fig. 2 Characteristic functions of typical voice, data and multimedia traffic in system with $\gamma^* = 1.55$ and $E_s^* = 0.3425$.

[†]It can be verified that for most S-shaped and increasing concave utility functions, the characteristic function will hold the single-peak shape.

in a distributed manner. In this way, the computational complexity is greatly reduced, and hence the scheme is practical for applications in real systems.

4. Algorithm Description and Implementation Discussion

4.1 Detailed Algorithm Processes

Assume the following system parameters are given when the system is set up: the base station transmitted power constraint $P_{\max}$, optimal SIR γ^* and corresponding efficiency E_s^*. The forms of characteristic functions $f_{\lambda i}(\cdot)$, together with the corresponding $r_{i\min}$ and $\lambda_{i\max}$ of each types of traffic should also be calculated in advance and stored in the radio resource allocation module. The following steps describe in detail how the proposed algorithms FCA and UCA work periodically at each radio resource allocation working point.

1. **[Adaption setup]**
   For all $1 \le i \le N$, current channel gain h_i and background interference I_i are predicted or measured at user terminals. The channel condition factor δ_i is calculated and returned to base station. Then initialize $\bar{\lambda}$ and $\underline{\lambda}$ to the maximal and minimal possible price, respectively, and set $\lambda = 0.382(\bar{\lambda} + \underline{\lambda})$.
2. **[Rate adaption]**
   For FCA, calculate R_i by Eq. (35); for UCA, calculate R_i by Eq. (34). The calculation is per-user based, therefore results can be obtained in a parallel, distributed manner.
3. **[Price adaption]**
   Adjust $\bar{\lambda}$, $\underline{\lambda}$ and λ by the golden section search: if $\sum_i \delta_i g(R_i) < 1$, $\bar{\lambda} = \lambda$, and $\lambda = 0.382(\bar{\lambda} + \underline{\lambda})$; otherwise $\underline{\lambda} = \lambda$, and set $\lambda = 0.618(\bar{\lambda} + \underline{\lambda})$.
4. **[Terminate condition]**
   If $\bar{\lambda} - \underline{\lambda} \le \varepsilon$, stop; otherwise goto step 2 for M times. Here ε is the required precision [18], and M is the maximal iteration times.
5. **[Normalization of resource allocation]**
   Calculate the normalized rate R_i by
   $$R_i^* = g^{-1}\left(\frac{g(R_i)}{\sum_j \delta_j g(R_j)}\right) \tag{36}$$
   where $g^{-1}(\cdot)$ is the inverse function of $g(R_i)$. By normalization, we get $\sum_{i=1}^{N} \delta_i g(R_i^*) = 1$ is satisfied, which means that the system resources are fully utilized.
6. **[Transmitted power]**
   Transmitted power assigned to user i is given by
   $$P_i = g(R_i^*)(P_{\max} + I_i^o) \tag{37}$$

As described above, the social optimal (sub-optimal for FCA) resource allocation can be obtained in a distributed manner with the adaption of resource price λ. Therefore, the computational complexity is significantly reduced compared with normal optimization algorithms, such as steepest decent method or gradient projection method [18], which may need hundreds of iterations to converge.

4.2 Implementation Issue

The utility-based resource allocation scheme proposed in this paper is highly applicable to real systems. Take WCDMA standard as a example. As we know, power control in WCDMA systems contains two hierarchical functional module: inner-loop power control (ILPC) and outer-loop power control (OLPC) [1], which are shown in Fig. 3. ILPC adjusts the transmitted power very fast (1500 times per second) to counteract fast channel fading. OLPC determines the relatively long-term transmission parameter (e.g., average transmitted power or target SIR) according to current system load and large time-scale channel conditions. It operates in time-scales much slower than ILPC (typically 10-50Hz in WCDMA systems). Actually, the resource allocation mechanism contains the OLPC module, and carries out the resource optimization process in the same frequency as OLPC.

The implementation of the utility-based resource allocation framework are illustrated in Fig. 4, only with some

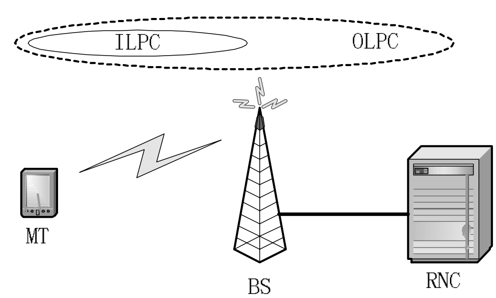

Fig. 3 Power control in WCDMA systems.

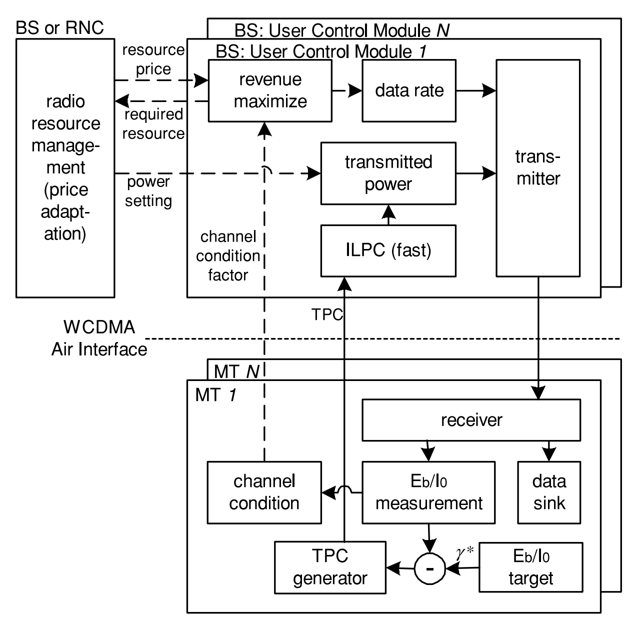

Fig. 4 Implementation structure of proposed utility-based radio resource allocation scheme and algorithms.

small modifications to the standard WCDMA ILPC/OLPC structure. As we see, there is a radio resource allocation module located in either base station or radio network controller (RNC) [1], which exchanges information with N user control modules and adjusts the resource price. User control modules calculate the required resources that maximizing users revenues. Also to implement proposed algorithms, some signalling or information flows should be added to the standard WCDMA structure, as the dashed lines shown in the figure. We see that only channel condition factors (δ_i) are needed to be added over air interface. Since the resource allocation works relatively slow, the information delivery overhead is totally acceptable.

With this implementation structure, the required resources are calculated in N user control modules parallelly, so that social optimal resource allocation is achieved in a distributed manner with lower computational complexity and shorter run time.

5. Numerical Results

We simulate a typical cell in DS-CDMA systems with the radius of 1km. A base station with omnidirectional antenna is located at the center of the cell. Tens of mobile users move in the cell with uniformly distributed velocities from 1 m/s to 30 m/s. Their initial locations and moving directions are randomly generated independently. Each mobile user alters its directions randomly after a exponentially distributed period of time. Since we do not focus on the handover issue, when a user reaches the boundary of the cell, it is bounced back.

Each user is equipped with one type of traffic among the four considered types: voice, data, mmedia1 or mmedia2. Their utility functions and characteristic functions are shown in Figs. 1 and 2, respectively. Parameters of these traffic are calculated and shown in Table 1. Other system parameters are assumed to be: $W = 5\,\text{MHz}$, $P_{\max} = 15\,\text{W}$, $\theta_d = 0.4$, and $I_i \sim (\mu, \sigma^2)$ with $\mu = 1 \times 10^{-12}\,\text{W}$ and $\sigma = 2 \times 10^{-13}\,\text{W}$. QPSK modulation, $(511, 175, 46)$ BCH coding (coding rate 1/3) and simple ARQ scheme are used as the physical layer technologies. According to [16], the transmission efficiency function is

$$E_s(\gamma) = \frac{175}{511}\left(1 - p_e Q\left(\frac{46 - 511p_e}{\sqrt{511p_e}}\right)\right) \tag{38}$$

where $Q(\cdot)$ is the residual distribution function of standard normal distribution $N(0, 1)$, and p_e is the raw bit error rate of block coding, which is a function of received SIR γ, presented by [15]

$$p_e = Q(\sqrt{2\gamma}), \quad \text{(For QPSK modulation)} \tag{39}$$

Table 1 Traffic parameters in simulation system.

Traffic Type	voice	data	mmedia1	mmedia2
$U_{i\max}$	1.6	8	5	15
$r_{i\min}$(Kbps)	46.72	0	187.3	1137.6
$\lambda_{i\max}$	3353	110.5	361.5	201.9

By numerical calculation we easily get that $\gamma^* = 1.55$ (1.9 dB) and $E_s^* = 0.3425$.

To shed light on the resource allocation algorithm, we ignore the working process of ILPC and assume that the small-scale fast channel fading are totally compensated by ILPC. Thus, only long-term channel condition is considered, and the mobile channel is modelled as

$$h_p = \overline{h_p} + h_s \tag{40}$$

where $\overline{h_p}$ is the long-distance propagation loss, given in dBs as [19]

$$\overline{h_p}(d) = -128.1 + 10n\log_{10}(d) \tag{41}$$

where d is the distance from user terminal to base station in Km; n is the loss index. According to [19], $n = 3.76$ for typical urban areas. The log-normal shadowing h_s is modelled as a stationary Gaussian stochastic process with the autocorrelation function of [20]

$$\begin{aligned} R_{h_s}(m) &= \mathrm{E}\{h_s(k)h_s(k+m)\} \\ &= \sigma_0^2 \eta_D^{|m|vT_s/D} \end{aligned} \tag{42}$$

where k is the sequence index, σ_0^2 is the variance of the log-normal shadowing in dBs, η_D is the correlation index between two points separated by a distance of D, v is the moving velocity of user terminals, and T_s is the time interval between two consecutive measurement points. Specifically we assume $\sigma_0 = 4.3\,\text{dB}$ corresponding to $\epsilon_d = 0.3$ and $D = 10$, as reported in [20].

We simulate 50 seconds in real time. The radio resource allocation module works 10 times per second (the time interval between two consecutive resource control work point is 100 ms). In order to analyze the performance, we also consider two other proposals given in [11] and compare them with our algorithm. One proposal is a resource control algorithm of congestion pricing for downlink (CPD), which is formulated as

$$\max_{P_i, R_i} U_i(R_i E_s(\gamma_i)) - \lambda P_i \tag{43}$$

We see that CPD is a per-user based optimization problem and it involves power and data rate joint control, which has rather high complexity. Another algorithm is called willingness-to-pay (WTP), with which the transmitted power is allocated by

$$P_i = \frac{w_i}{\sum_j w_j} P_{\max} \tag{44}$$

where w_i is the user i's paying willingness [11]. Here we use the value of $r_{i\min}$ to represent it.

In Fig. 5 we compare the system overall utility of of CPD, WTP and our algorithms. In the figure x-axis denotes 5 different simulation scenario with the simulation environments (user location, movements, traffic type, etc.) being generated randomly. Figure 6 shows the comparison of

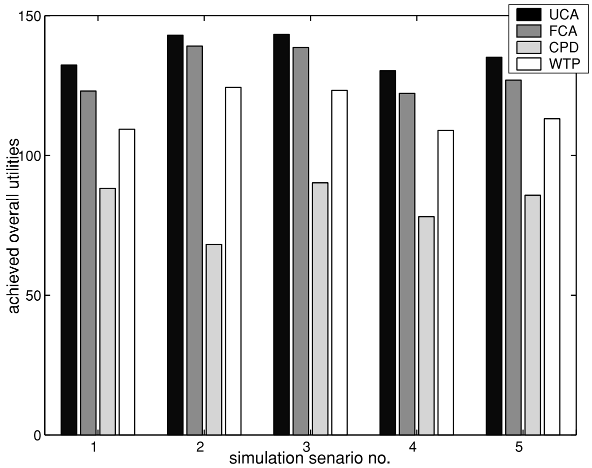

Fig. 5 Comparison of achieved system overall utilities: UCA, FCA, CPD and WTP.

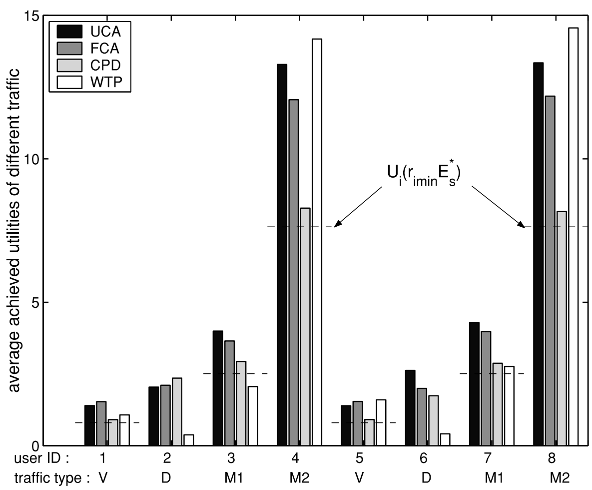

Fig. 6 Comparison of average achieved utilities of some typical users with different traffic types: UCA, FCA, CPD and WTP.

achieved utilities of typical users in a typical simulation scenario. In Fig. 6, the value of $U_i(r_{i\,\min}E_s^*)$, which is the corresponding utility of $r_{i\,\min}$, is drawn. From the two figures we clearly observe that:

- UCA and FCA both achieve more overall utilities than CPD and WTP, as shown in Fig. 5. Especially compared with CPD, the utility gain is generally up to 50–90%, which is significant. Though WTP performs better than CPD, it still achieves less overall utilities than our proposal.
- FCA obtains less overall utilities than UCA, which verifies the analysis in Sect. 3 that FCA get suboptimal resource allocation.
- Fairness is well guaranteed by UCA and FCA. As shown in Fig. 6, for all proposals, the average achieved utilities of all types of traffic exceed $U_i(r_{i\,\min}E_s^*)$. However, for WTP, data users suffer "service starvation" in that their utility level is very low though other users keeps relatively high. Hence, it is totally unfair for data users.

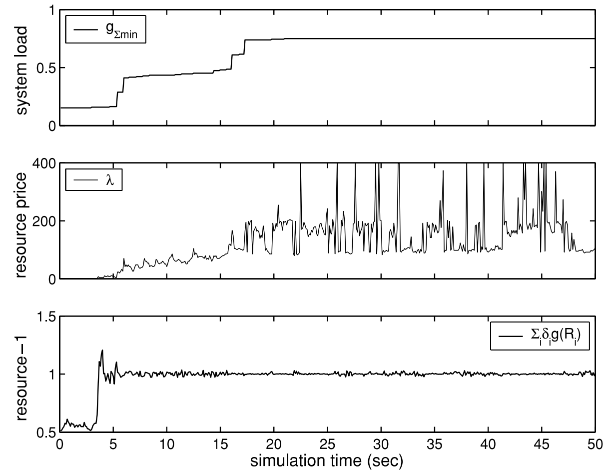

Fig. 7 Adaption and convergence observation of resource price λ in FCA algorithm.

As a conclusion, we see that UCA and FCA have advantages over CPD and WTP in overall utilities and better user fairness, therefore our proposal are more appropriate and efficient for radio resource allocation in DS-CDMA systems.

Then we take a detailed look on how our algorithms work in a typical simulation scenario. First the dynamic behavior of the network price λ corresponding to the variation of system load is depicted in Fig. 7. We take FCA as an example. In the figure, the system load is represented by $\sum_{i=1}^{N} g(r_{i\,\min})$, which is designed to be with an increasing trend. From the figure we observe:

- The resource price fluctuates drastically. From the system model we know that, the time-varying channel conditions will influence resource price. When channel conditions are relatively good (the channel condition factor δ_i decreases), less resource is need to be consumed to counteract the background interferences. From the market's point of view, the demands of the commodity decreases. As a result, resource price λ drops so that users are willing to consume more resources. On the other hand, when the wireless link conditions are deteriorated, resource price rises to prevent users to over-consume radio resources.
- When the system load increases, each user has to consume less resources so that the system can provide service to all users simultaneously. Therefore, in a heavily loaded system, λ also has a trend to increase.

Also the resource allocation before the normalization is shown in Fig. 7 ("resource-1" in the figure, which means $\sum_{i=1}^{N} \delta_i g(R_i)$ before the step of normalization, see Sect. 4). It clearly indicates that the algorithm adapts itself to the time-varying mobile channels. From this figure we see that the maximal iteration time $M = 6$ is good enough for the algorithm to converge to the optimal resource allocation. For UCA algorithm, the figures are slightly different, but similar

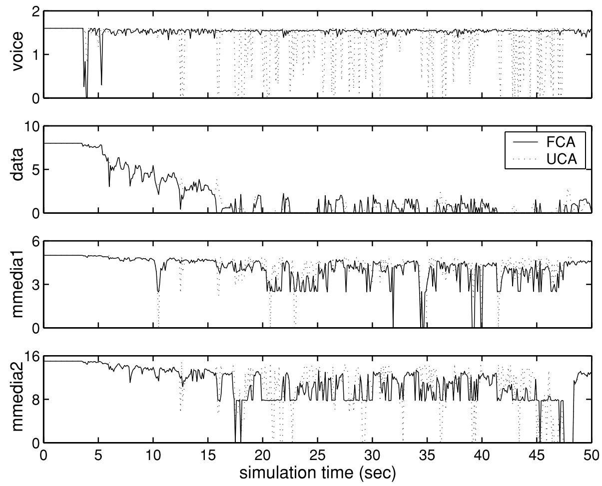

Fig. 8 Simulation trace of the achieved utilities of voice, data and multimedia users in a typical simulation scenario.

conclusions can be drawn.

Next in Fig. 8 we present the simulation trace of the four types of traffic in the system. We see that the for voice traffic, the achieved utility keeps almost unchanged at different system load; though the achieved utilities of the two types of multimedia traffic are influenced by the variation of system load, they also keep at a relatively high level. On the other hand, traffic load has strong influence on the achieved utility of data traffic: when system is heavily loaded, data traffic withdraws from the contention of resources automatically, and system radio resource is saved for traffic with higher priority, i.e., data and multimedia traffic. Another phenomenon worth mentioning is that: compared with FCA, users of UCA experience more service starvation. For example, for the observed mmedia2 user, it is cut off much more often when the system load is high, while FCA allocates minimal resource to the same user in most cases. This is very helpful for the performance of application layer protocols, since frequent dropping-to-zero of data rate may leads to the case of time-out and lost of connections in upper layer protocols and applications. In other words, compared with UCA, FCA obtains better fairness in the expense of a small amount of overall utilities.

6. Conclusions and Future Work

In this paper, we presented an utility-based framework for radio resource allocation problems in the downlink of DS-CDMA systems carrying multimedia traffic, where transmitted power and data rate are controllable resources. The goal is to achieve the social optimal allocation, which is the maximization of system overall utilities while the transmitted power constraint of base station being satisfied. This utility-based radio resource allocation problem is formulated as an optimization model mathematically.

In order to avoid the high computational complexity of solving the nonlinear optimization problem, we reformulated it into an market model, where radio resource is regarded as a commodity. It is supplied by the network, and is consumed by users. In the market model, two subproblems are involved: one is revenue maximization subproblem, which finds the optimal transmission data rate of each individual user according to their own utility function, channel condition, and current resource price. The other subproblem is market equilibrium. By dynamic adaption of resource price, the market converges to the equilibrium, where supplies of resource equals demands. Since the market model satisfies the incentive-compatibility constraints, the social optimal allocation is achieved at the market equilibrium. In this way, the resource allocation problem is solved in a distributed manner, and the computational complexity is significantly reduced. The radio resource allocation framework can be easily applied in WCDMA systems, and we also presented the implementation structure.

According to whether to concern the fairness issue of users, two different utility-based radio resource allocation algorithms, UCA and FCA, were proposed. UCA, simply cutting off the user when the resource price is too high, focus on the achieving the maximal overall utilities, while FCA provides minimal transmission data rate guarantees to users that cannot afford to consume the resource.

It can be observed from simulation results that both UCA and FCA are flexible and efficient for the downlink of DS-CDMA systems. Compared with UCA, FCA obtains better fairness in the expense of a small portion of overall utilities. We also compared the performance of proposed resource allocation algorithms with other ones described in previous researches. Simulation results have been shown that the UCA and FCA have advantages on the overall performance (system overall utilities) as well as the fairness for users over other schemes. Therefore, they are feasible for radio resource allocation in multimedia DS-CDMA systems.

We do not consider the effect of delay to the utility functions in this paper, which may be important to some delay-sensitive traffic. It is a part of the on-going work. In addition, we will analyze the influence of channel estimation or prediction errors to the performance of the proposed algorithms in future study.

Acknowledgment

The authors are grateful for Dr. David Lee, Dr. Dawei Huang and other researchers of Bell Labs Research China (BLRC), Lucent Technologies, for fruitful discussions and valuable suggestions.

References

[1] S. Dixit, Y. Guo, and Z. Antoniou, "Resource management and quality of service in third-generation wireless networks," IEEE Commun. Mag., vol.39, no.2, pp.123–133, Feb. 2001.
[2] A. Sampath, P.S. Kumar, and J.M. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," Proc. IEEE PIMRC95, vol.1, pp.21–25, 1995.
[3] M.A. Arad and A. Leon-Garcia, "Scheduled CDMA: A hybrid multiple access for wireless ATM networks," Proc. IEEE PIMRC96, Taipei, Taiwan, 1996.

[4] S. Shenker, "Fundamental design issues for the future Internet," IEEE J. Sel. Areas Commun., vol.13, no.7, pp.1176–1188, 1995.

[5] M. Xiao, N.B. Shroff, and E. Chong, "Utility-based power control (UBPC) in cellular wireless systems," Proc. IEEE INFOCOM2001, vol.1, pp.412–421, 2001.

[6] J.W. Lee, R.R. Mazumdar, and N.B. Shroff, "Downlink power allocation for multi-class DS-CDMA wireless networks," Proc. IEEE INFOCOM2002, http://www.ieee-infocom.org/2002/papers/309.pdf

[7] V. Shah, N.B. Mandayam, and D.J. Goodman, "Power control for wireless data base on utility and pricing," Proc. IEEE PIMRC98, vol.3, pp.1427–1432, 1998.

[8] D.J. Goodman and N.B. Mandayam, "Power control for wireless data," IEEE Pers. Commun., vol.7, no.2, pp.48–54, Feb. 2000.

[9] H. Ji and C.Y. Huang, "Non-cooperative uplink power control in cellular radio systems," ACM/Baltzer Wireless Networks Journal (WINET), vol.4, pp.233–240, April 1998.

[10] T. Alpcan, T. Basar, R. Srikant, and E. Atman, "CDMA uplink power control as a noncooperative game," Proc. 40th IEEE Conf. Decision and Control 2001, vol.1, pp.197–202, 2001.

[11] V.A. Siris, "Resource control for elastic traffic in CDMA networks," Proc. 8th Annual Int. Conf. Mobile Computing and Networking (Mobicom2002), pp.193–204, Oct. 2002.

[12] P. Marbach and R. Berry, "Downlink resource allocation and pricing for wireless networks," Proc. IEEE INFOCOM2002, vol.3, pp.1470–1479, June 2002.

[13] H.R. Varian, Intermediate Microeconomics: A Modern Approach, translated into Chinese, W.W. Norton & Co., New York, London, 1990.

[14] C.-L. I and K.K. Sabnani, "Variable spreading gain CDMA with adaptive control for true packet switching wireless network," Proc. IEEE ICC95, vol.2, pp.725–730, June 1995.

[15] T.S. Rappaport, Wireless Communications: Principles and Practice, Prentice Hall, Simon & Schuster, 1996.

[16] X. Duan, Z. Niu, and J. Zheng, "Analysis of wireless transmission efficiency and its application: Efficiency-based adaptive coding," Proc. IEEE PIMRC2003, Beijing, China, Sept. 2003.

[17] X. Duan, Z. Niu, and J. Zheng, "Capacity analysis of uplink and downlink in multimedia DS-CDMA systems based on constraint models," Proc. IEEE ICC2003, vol.1, pp.402–406, Alaska, USA, May 2003.

[18] Authors Committee of Modern Applied Mathematical Manual book series, Modern Applied Mathematical Manual: Volume on Operational Research and Optimization Theory, Tsinghua Univ. Publishing House, Beijing, 1998.

[19] 3GPP TR 25.942 (Release 4), "RF System Scenarios," 3GPP Technical Specification Group Radio Access Networks, 2001.

[20] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," Electron. Lett., vol.27, pp.2145–2146, Dec. 1991.

Xiang Duan received his B.S. degree in the Department of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, China, in 1999. Currently, he is a Ph.D. candidate in the Department of Electronic Engineering, Tsinghua University. His research interests include radio resource management, quality of service control in wireless networks, wireless and mobile applications, and ad hoc networks.

Zhisheng Niu graduated from Northern Jiaotong University, Beijing, China, in 1985, and got his M.E. and D.E. degrees from Toyohashi University of Technology, Toyohashi, Japan, in 1989 and 1992, respectively. In 1994, he joined with Tsinghua University, Beijing, China, where he is now a full professor at the Department of Electronic Engineering. He is also an adjunct professor of Northern Jiaotong University. From April 1992 to March 1994, he was with Fujitsu Laboratories Ltd., Kawasaki, Japan. From October 1995 to February 1996, he was a visiting research fellow of the Communications Research Laboratory of the Ministry of Posts and Telecommunications of Japan. From February 1997 to February 1998, he was a visiting senior researcher of Central Research Laboratory, Hitachi Ltd. From January 2001 to February 2001, he was a visiting professor of Venture Business Laboratory at Saga University, Japan. In February 2002, he also visited Polytechnic University, New York, USA. He received the PAACS Friendship Award from the Institute of Electronics, Information, and Communication Engineers (IEICE) of Japan in 1991 and the Best Paper Award (1st prize) from the 6th Chinese Youth Conference on Communication Technology in 1999. His current research interests include performance evaluation of high-speed broadband integrated networks, wireless/mobile IP networks, resource management and traffic control, and High Altitude Stratospheric Platforms. Dr. Niu is a senior member of the IEEE, IEEE Communication Society Beijing Chapter Chair, Membership Development Committee Chair of IEEE ComSoc Asia-Pacific Board, and Chair of APCC Steering Committee. He is also a senior member of the Chinese Institute of Electronics (CIE), a member of the Chinese Society of Operation Research.

Junli Zheng graduated from Tsinghua University, Beijing, China, in 1961. Since 1961, he has been with the Department of Electronic Engineering, Tsinghua University, Beijing, China, where he is now a professor. He is a member of China Institute of Circuits and Systems and China Neural Network Committee. His current research interests include signal processing for communications, neural network, chaos theory and their applications.